Technical Specification Group Services and System Aspects *TSGS#14(01)0680*
Meeting #14, Kyoto, Japan, 17-20 December 2001

Source:	**SA1**
Title:	**TS 22.243 v1.0.0 on Distributed Speech Recognition**
Document for:	**Information**
Agenda Item:	**7.1.3**

TSG-SA WG 1 (Services) meeting #14 **S1-011313**
Kobe, Japan, 5-9 November 2001 **Agenda Item: 9.8**

Source: DSR ad-hoc chair

Presentation of Specification to TSG or WG

Presentation to:	**TSG SA#14, SA WG2, SA WG3, SA WG4, SA T2, CN WG1**
Document for presentation:	**TS 22.243, Version 1.0.0**
Presented for:	**Information**

Abstract of document:

Stage 1 description for inclusion of ETSI Aurora based Distributed Speech Recognition as a 3G service.

Changes since last presentation

Not previously presented.

Outstanding Issues:

–Should Aurora be an IMS service only or should it also be a requirement to support on CS domain?

–Clarify Codec negotiation requirements.

–Be consistent in use of “user” and “subscriber”.

–QoS requirements and classes for DSR?

–Service example needed to clarify requirements of third parties.

–Is there a privacy requirement?

–Some definitions are required.

Completion of the above Expected in February 2002.

Contentious Issues:

None

3GPP TS 22.243 V1.0.0 (2001-11)

Technical Specification

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
Distributed Speech Recognition Based Automated Voice;
Services Stage 1
(Release 5)

Keywords

<keyword[, keyword]>

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

Contents

Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Forecasts show that speech-driven services will play an important role on the 3G market. People want the ability to access information while on the move and the small portable mobile devices that will be used to access this information need improved user interfaces using speech input. At present, however, the complexity of medium and large vocabulary speech recognition systems are beyond the memory and computational resources of such devices.

Server-side processing of the combined speech and DTMF input and speech output, can overcome these constraints by taking full advantage of memory and processing power as well as specialized speech engines and data files. However, the distortions introduced by the encoding used to send the audio between the client and the server as well as additional network errors can significantly degrade the performance of the speech engines; therefore also limiting the achievable speech functionalities. A server-side speech service is generally equivalent to a phone call to an automatic service. As for any other telephony service, DTMF is a feature that should always be considered as needed.

In this document, Distributed Speech Recognition (DSR), is a generic framework to distribute the audio sub-system and the speech services by streaming encoded speech between the client and the server. Instead using a speech channel as in today's server-based speech services, an error-protected data channel will be used to transport the speech code from the front-end (terminal client) to the speech engine (on server). The additional use of appropriate DSR codecs optimised for speech recognition overcomes the problems identified above.

DSR will provide users with a high performance distributed speech interface to server based automatic services for communication, information or transactional purposes.

The types of user interfaces include those that are voice only, for example, automatic speech access to information, such as a voice portal described in this section, including combined speech / DTMF input.

In the future, a new range of multi-modal applications is also envisaged incorporating different modes of input (e.g. speech, keyboard, pen) and speech and visual output.

ETSI STQ Aurora has developed a DSR framework that attempts to minimize the negative effect of encoding and network losses on the speech engine performance. This includes:

- transport protocols (IETF AVT DSR), codec negotiation / switch and meta-information exchanges.

- a DSR optimised codec (ETSI ES 201 108 [4]) and studied the performance improvements that it provides over other (e.g. AMR) codecs. This encoding scheme is supported by major speech recognition and handset manufacturers.

This framework and the DSR optimised codecs can be taken as starting points.

1 Scope

This Technical Specification defines the stage one description of the Distributed Speech Recognition Based Automated Voice Service. Stage one is the set of requirements for data seen primarily from the subscriber's and service providers' points of view.

This TS includes information applicable to network operators, service providers, terminal and network manufacturers.

This TS contains the core requirements for the Distributed Speech Recognition Based Automated Voice Service.

The scope of this Stage 1 is the ETSI Aurora solution adapted for 3GPP.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TS 21.133: "3G Security; Security Threats and Requirements"

[2] 3GPP TS 21.905: "Vocabulary for 3GPP Specifications"

[3] 3GPP TR 22.941: " IP Based Multimedia Services Framework"

[4] ETSI ES 201 108 v1.1.2 Distributed Speech Recognition: Front-end Feature Extraction Algorithm; Compression Algorithm", April 2000.

3 Definitions and abbreviations

Definitions and abbreviations used in the present document are listed in TR 21.905 [2]. For the purposes of this document the following definitions and abbreviations apply:

3.1 Definitions

Automated Voice Services:

Distributed Speech Recognition:

Distributed Speech Recognition Based Automated Voice Service:

Distributed speech recognition codec:

Text-to-Speech Synthesis:

Meta information:

3.2 Abbreviations

For the purposes of this document the following abbreviations apply:

DSR – Distributed Speech Recognition.

IMS – IP Multimedia Subsystem

URI – Universal Resource Identifier

4 Requirements

The following list gives the high level requirements for the Distributed Speech Recognition Based Automated Voice Service:

Subscribers to the Distributed Speech Recognition Based Automated Voice Service shall be able to setup voice communication, access information or conduct transactions by voice commands using distributed speech recognition (DSR). Example Automated Voice Services include:

- communication assistance (Name dialing, Service Portal, Directory assistance)
- Information retrieval (e.g., obtaining stock-quotes, checking local weather reports, flight schedules, movie/concert show times and locations)
- M-Commerce and other transactions (e.g., buying movie/concert tickets, stock trades, banking transactions)
- Personal Information Manager (PIM) functions (e.g., making/checking appointments, managing contacts list, address book, etc.)
- Messaging (IM, unified messaging, etc…)
- Information capture (e.g. dictation of short memos)

For the sake of brevity, these services will be referred to as ‘DSR-based Automated Voice Service’ in the rest of this section.

The DSR-based Automated Voice service will be offered by the network operators and will bring value to the network operator by the ability to charge for the automated voice services.

This service shall be offered over the IMS (editor’s note - or circuit switched also). The protocols used for the uplink streaming of DSR front-end parameters (from terminal to server) and associated control and application specific information will be based on those in IMS.

4.1 Initiation

It shall be possible for a user to initiate a connection to the DSR-based Automatic Voice Service, for example, by entering the identity of the service. The identity used will depend on the scheme of the service provider but could include a phone number or even a URI.

Editor’s note. May need to be verified depending on the supported domain.

4.2 Information during the call

At initiation and potentially during the call it will be possible to select the uplink and downlink codecs. The terminal shall support a standard DSR optimized codec as default uplink codec. When supported by the terminal and the automated voice service, another uplink codec can be selected. This may be motivated by the expected or observed acoustic environment, the service package purchased by the user, the user profile (e.g. hands-free as default) or service need. The downlink codec shall be negotiated as conventionally done within 3GPP.

Editor's note - Uplink negotiated?

The user speaks to the service and receives output back from the automated voice service provider as audio (recorded 'natural' speech or Text-to-Speech Synthesis. Additional control and application specific information shall be exchanged during the call between the client and the service. Accordingly some terminals shall support sending additional data to the service (e.g. keypad information and other terminal events) and receiving data feedback that shall be displayed on the terminal screen.

4.3 Control

It shall be possible for users to access and navigate within and between the various automated speech services by spoken commands or pressed keypads. Simultaneous uplink / downlink transfer shall be supported.

It shall be possible for network operators to control access to services based on subscription profile of the callers.

Editor's note - User or subscriber?

4.4 User Perspective (User Interface)

The user's interface to this service shall be via the terminal equipment. User can interact by spoken and keypad inputs. The terminal equipment can have a visual display capability. When supported by the terminal, the server-based application can display visual information (e.g., stock quote figures, flight gates and times) in addition to audio playback (via recorded speech or text-to-speech synthesis) of the information.

5 Service Specific Considerations

5.1 Authorization

Authorization for use of this service shall be under the control of the network operator.

5.2 Deauthorization

Deauthorization for use of this service shall be under the control of the network operator.

5.3 Registration

Connection to the IMS with terminal equipment containing the requirement capabilities to allow the use of the DSR-based Automated Voice service shall be possible.

5.4 Deregistration

Disconnection from the IMS shall prevent the use of the DSR-based Automated Voice service.

5.5 Activation

Activation shall be as for IMS Basic Voice Service. For further study

5.6 Deactivation

Deactivation shall be as for IMS Basic Voice Service. For further study

Editor's note 5.3 to 5.6 are subject to decision regarding support of domains.

6 Service Provisioning

The DSR-based Automated Voice Service shall be able to be provisioned by either the network operator or by a 3rd party service provider.

6.1 Distribution, downloading, terminal capabilities

In addition to the basic capabilities for voice (such as the default voice codec that will be used for the downlink DSR audio prompt stream), the following DSR-specific capabilities shall be required in the UE:

- A standard default uplink optimised DSR codec.
- A downlink conventional codec (simultaneous with uplink)
- The capability to transmit keypad information from the client to the server (e.g., either DTMF or the keypad string)
- Application specific information exchanges shall be supported between the client and the server (client events, display information etc…)

Editor's note - Clarify what information, and is it implementation specific?

6.2 Service Provisioning Requirements

In addition to the requirements for the Basic Voice service, the following capabilities shall be provided for this service: *** some issues are duplicated from 6.1 ***

- The terminal equipment shall have the capability to transmit keypad information to the server (e.g., either DTMF or the keypad string)
- Application specific information exchanges shall be supported between the client and the server (client events, display information etc…)
- The terminal equipment shall have at least a standardized default uplink DSR codec.
- The network shall have uplink streaming capability to enable transmission of audio input from the terminal to the automated voice service on the server.
- DSR shall be supported by network QoS (Quality of Service) for conversational class services. For example, 150 msec expected and 400 msec maximum for end-to-end one-way delay.

Editor's note Restricted to conversational class? What are the QoS requirements on DSR?

7 Security

The “Security Threats and Requirements” specified in 21.133 [1] shall not be compromised.

It shall be possible to deny unauthorized access to the 3GPP Distributed Speech Recognition Based automated Voice Service. An authorization may be based on the following,

- identity of the accessing user agent, server or device
- the destination user, device or user agent

Third parties shall have authorization from the User and PLMN Operators in order to access the 3GPP Distributed Speech Recognition Based automated Voice Service.

Editor’s note - Example needed

8 Privacy

It shall be possible for the user to express privacy requirements for the Distributed Speech Recognition Based automated Voice Service that can be used to determine access rights.

Editor’s note - Justification of this requirement is needed.

9 Charging

The user can be charged for sessions with the DSR-based automated Voice Service in a variety of ways. The following shall be possible:

a) By duration of call (including “one-off” charge/flat rate)

b) By data volume transferred (number of packets)

c) By subscription fees for the service (unlimited usage or unlimited usage up to a point and then per-use fees)

d) Free (with the service being subsidised by advertising revenue from advertisement spots)

The advertisement spots may be inserted either at session start-up or close, or designed in such that system delay time is masked (e.g., while the user is waiting for the flight schedules to be returned, or a purchase transaction to be completed).

The network operator will receive revenue from subscribers directly as well as from the content and service providers who want their sites to be accessible via the automated voice service, and from advertisers. Advertising spots can be inserted at appropriate points during the session (e.g., at the beginning of the session, while the user is waiting for a system response, or at the end of a session).

Editor’s note – Need to distinguish DSR from a basic IMS voice call. Charge according to QoS subscribed.

10 Roaming

The subscriber shall be able to utilize the DSR-based automated Voice Service when roaming in any IMS compatible mobile network, where there is an appropriate roaming agreement established. The DSR parameters could be routed back to the home network, if delays are appropriately managed. However, network latencies introduced by traversal over multiple networks might necessitate (seamless) switchover to a closer service provider. For this reason, the capabilities of the DSR-based automated Voice Service shall be available in the roamed-to network in the same manner as in the home network; within the limitation of the capabilities of the serving network.

Editor's notes – Subject to decision on domain supported. Implementation by providing in visited network. Delete after first sentence?

11 Interaction with other services

It shall be possible for the DSR-based automated Voice Service to be combined with the Basic Voice service. This way, all users of the DSR-based automated Voice Service should be able to use automated voice services on any telecommunications network

Editor's note - Why limit only to interaction with Basic Voice Services? Interact with other service if other domains used.

Annex A (informative): Change History

Change history											
TSG SA#	**SA Doc.**	**SA1 Doc**	**Spec**	**CR**	**Rev**	**Rel**	**Cat**	**Subject/Comment**	**Old**	**New**	**Work Item**
			22.xxx					Initial draft based on content of TR 22.941		0.0.1	
			22.xxx					Output of ad-hoc drafting session, Kobe, 11-07-01	0.0.1	0.0.2	
			22.243					Raised to version 1.0.0 by SA1#14 for presentation to TSG-SA#14 for information	0.0.2	1.0.0	

SA1 Rapporteur: David Williams, QUALCOMM, Inc.

Email: dwilliams@qualcomm.com Telephone: +33 6 72019738